



**Project no.:** FP7-ICT-217077  
**Project full title:** Heterogeneous 3-D Perception across Visual Fragments  
**Project Acronym:** EYESHOTS  
**Deliverable no:** D4.3a  
**Title of the deliverable:** How to build a global awareness of the peripersonal space

<b>Date of Delivery:</b>	08 September 2010	
<b>Organization name of lead contractor for this deliverable:</b>	UJI	
<b>Author(s):</b>	E. Chinellato, M. Antonelli, J.B. Grzyb, A.P. del Pobil	
<b>Participant(s):</b>	UJI, WWU	
<b>Workpackage contributing to the deliverable:</b>	WP4	
<b>Nature:</b>	Report	
<b>Version:</b>	1.1	
<b>Total number of pages:</b>	11	
<b>Responsible person:</b>	Angel P. del Pobil	
<b>Revised by:</b>	Markus Lappe, Svenja Wulff	
<b>Start date of project:</b>	1 March 2008	<b>Duration:</b> 36 months

<b>Project Co-funded by the European Commission within the Seventh Framework Programme</b>		
<b>Dissemination Level</b>		
<b>PU</b>	Public	<b>X</b>
<b>PP</b>	Restricted to other program participants (including the Commission Services)	
<b>RE</b>	Restricted to a group specified by the consortium (including the Commission Services)	
<b>CO</b>	Confidential, only for members of the consortium (including the Commission Services)	

**Abstract:**

This deliverable describes the extension of the model for sensorimotor awareness of the peripersonal environment to the 3D space, and its application on the UJI humanoid robot setup. Results of simulated psychophysical experiments regarding the saccadic adaptation paradigm are provided. Object recognition as it is possibly performed by the ventral stream is presented, and an application schema for integration with dorsal sensorimotor information is provided. Robotic implementation of the theoretical principles and related ongoing experimentation are described.

## Contents

1. Introduction.....	3
2. Dorsal stream: visuomotor representation of objects in the peripersonal space .....	3
3. Ventral stream: object recognition .....	5
4. Reproduction of psychophysical findings – saccadic adaptation .....	6
5. Stream integration – sensorimotor memory of nearby objects .....	9
6. Hardware setup and setting of robotic experiments. ....	9

## Executive summary

In this document we describe the full functionality, extended to the three-dimensional space, of the peripersonal space sensorimotor awareness model, described in D4.2b, and its implementation on the UJI robotic platform. Employing the Radial Basis Function framework developed in the model, the robot is able to achieve open loop gazing and reaching capabilities toward visual targets, according to the goals of Task 4.2.

According to milestone M12, also planned for month 30, the computational model of joint visuo-motor representation of the 3D space based on neurophysiological/psychophysical findings was completed and applied to the robotic setup. The system functionality according to the model is working under controlled situations and is now being tested extensively. The system is able to reach a limited set of artificial objects in a controlled real environment as the result of its perceptual awareness about an object and the modulation of the reaching and gaze directions.

## 1. Introduction

The model for building an integrated sensorimotor knowledge of the environment developed so far and described in Chinellato *et al.* 2010 has been suited to be applied on the UJI humanoid robot setup. The robotic implementation of such model includes: 1) a module for visual acquisition and visual processing that generates disparity information regarding simple visual stimuli, such as point-like features on plain backgrounds; 2) the RBF structure for the contextual representation of stimuli in multiple reference frames, extended to the 3D space as explained in Section 3 of this document; 3) modules for controlling the execution of saccadic eye movements by the robot stereo head and arm reaching movements.

As expected, basic skills such as concurrent or decoupled gazing and reaching movements toward simple visual stimuli are available to the robot. As a first experimental test, the system is able to show its visuomotor capabilities by performing oculomotor actions toward simple targets placed in its visual environment, or toward the location where its hand lies (identified with markers). Complementarily, it is also able to perform arm reaching movements to a similar visual target, either with or without gazing at it. The latter is a case of peripheral reaching, in which an intermediate transformation from visual to oculomotor space is performed, but the corresponding motor signal is not released.

As described in this document, the model emulates some psychophysical effects related to deceptive feedback, such as in the saccadic adaptation paradigm, and such effects are now being explored on the robotic implementation searching for further insights that could come from the application of the model on a real world setup. Extensive experimentation of such capabilities is now under development, in parallel with the integration with WPs 1, 2 and 3 modules, which are expected to improve and complement such abilities.

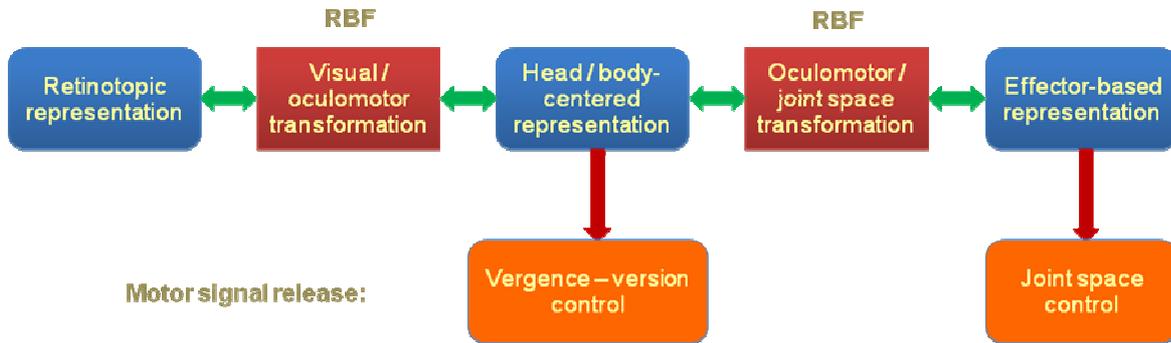
In Section 2 the model framework extended to the 3D space is presented, and some implementation details are provided. In Section 3 the ventral contribution to the visuomotor knowledge of the environment, in form of object recognition, is introduced. A description of saccadic adaptation experiments simulated with the model is provided in Section 4. Section 5 describes how dorsal and ventral contributions will fuse to achieve a purposeful interaction of the agent in a multi-object environment. Finally, robotic implementation with preliminary results and description of the ongoing experiments is given in Section 6.

## 2. Dorsal stream: visuomotor representation of objects in the peripersonal space

The modeling framework presented in this section is an extended version of the computational model described in deliverable 4.2b and the annexed paper (Chinellato *et al.* 2010). The most important development is that the model and the corresponding neural network structure have been extended to work in a real 3D environment. The main characteristics of the computational schema and its implementation, based on insights from WP5 experiments and especially V6A data, do not change in this version of the model. The framework depicted in Fig. 1, introduced in the previous deliverable referring to a 2D environment, is thus still fully representative of our modeling effort.

At this stage, the model is designed to deal with stimuli in a three dimensional environment, where the robot can gaze to and/or reach for an object in its peripersonal space. The model is composed of two parts. The first part is a radial basis function network which transforms the retinotopic position of the stimulus into the angular positions of the head motors that are necessary to gaze at the stimulus. The inputs of this network are the position of the stimulus on a cyclopean image ( $x, y$ ) and its horizontal disparity, while the outputs are the angular positions of the head motors, namely the vergence and the version ( $x, y$ ) of the eyes. The second part is responsible to recode oculomotor coordinates in arm joint space and vice-versa. It is made of two RBF networks that use the same parameters and the same training points to learn to reach with the hand for the fixation point (direct transformation) and to foveate the position of the hand (inverse transformation). The

oculomotor joint/space RBFs transform vergence and the version (x, y) inputs into angular positions of the shoulder (two d.o.f) and elbow (one d.o.f), and vice-versa.



**Figure 1:** Computational framework.

The transformation functions implemented by the RBF networks depend on some parameters that are implicitly learnt by the model. For the visual/oculomotor transformation, the parameters are interocular distance, focal length and resolution of the camera. Focal distortions were ignored in the simulated model, assuming that the application of the learning framework on the robot will allow to compensate for any known or unknown distortions and also adapt to changing parameters. The parameters included in the oculomotor/joint space RBFs are five: lengths of arm and forearm, interocular distance, and relative position of shoulder and eyes (two parameters, supposing they are aligned in the z coordinate).

The networks which compose the two parts of the model are radial basis function networks with three layers: input, hidden and output. The hidden layer performs the non-linear transformation of the input, and the output is computed as a weighted sum of the hidden neurons' activity. The parameters of the networks were experimentally chosen a-priori, so the learning phase affects only the weights which link hidden and output layers. For the first part we employed hidden neurons with Gaussian activation logarithmically distributed on the input space. The values of the radii were set to be proportional to the distance between adjacent centers.

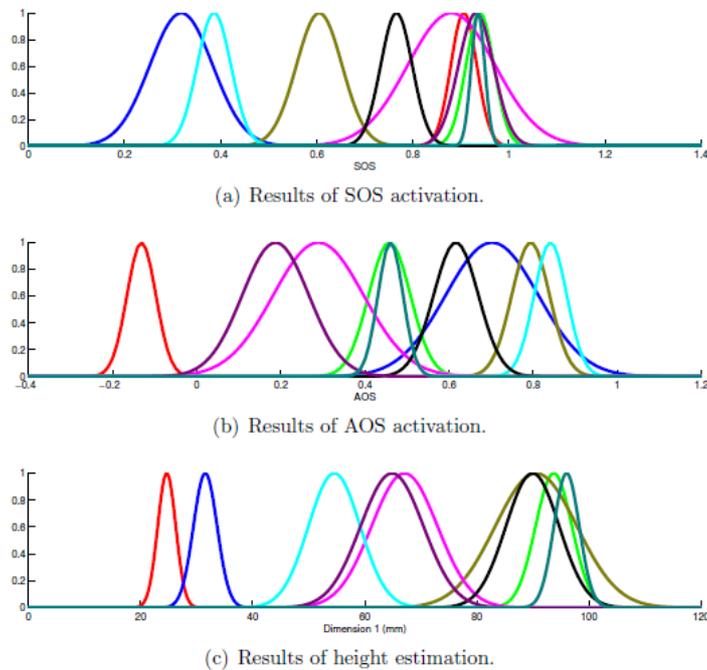
The direct and inverse transformations between head and arm motors were performed with centers uniformly distributed on the input space and having the same Gaussian activation. In the simulated model of the robot, the weights of the networks were calculated minimizing the mean square error over a given training set of points homogeneously distributed in a Cartesian space.

To pre-adapt the model for its application on the real robot, an on-line delta-rule learning strategy has been developed, which allows the robot to continuously learn its parameters and progressively improve its gazing and reaching capabilities. In order to estimate the sort of results we could expect applying the computational framework to the robotic setup, we changed the kinematic parameters of the robot model, and start training the network with the old weights from the new configuration. Considering for example the more demanding oculomotor to joint-space transformation, after modifying the above parameters about 10% to 20%, the error rises up immediately, and drops back almost to the original precision only after about 50 trials. This behavior shows the adaptability of the system to changes in working conditions, and supports its suitability for implementation on the robot, as described in Section 6.

### 3. Ventral stream: object recognition

In addition to building a sensorimotor map of visual and motor targets in the nearby space, the construction of an integrated knowledge of the environment requires also the identification of objects or targets and the use of a memory of previously observed/reached objects.

An important open issue that needs to be solved is how to associate object identity to motor memory of locations. So far, we have dealt with the object recognition task to identify simple shapes according to various available features. Among others, we include visual features regarding object proportion, a sort of information that is very likely forwarded to the ventral stream by dorsal areas. Such features are the activation of Surface Orientation Selective (SOS) and Axis Orientation Selective (AOS) Neurons of the posterior intraparietal sulcus area CIP (Chinellato and del Pobil, 2008). In addition, we include the estimated size of the object three main axes, its general shape (approximated to box, cylinder or sphere) and its color, for a total of seven features. The color feature is dominant, and no other features would be required if objects had different colors. In any case, we are interested in having a robust visual object identifier, able in principle to recognize objects of the same material and thus with very similar colors.



**Figure 2:** Discrimination of object identity with SOS and AOS data and estimated object height, for nine objects.

Our tests so far have been reduced to a box universe. At first, the system is provided with a number of labeled items, and uses visual perception to associate detected features to object identity. Then, upon presentation of a novel input, it tries to recognize the visualized object, through a probability density estimator. The system decides that a target object has been recognized when there are at least five features that match one of the known classes. A feature matches if the sample is in the range  $[\mu - n\sigma, \mu + n\sigma]$ , where parameter  $n$  defines the tolerance of the classifier, and has been set to 2, corresponding in theory to about a 95% probability of correct classification ( $n = 3$  would correspond to approximately 99%). If a sample is classified, it is directly added to the classifier memory to be used in subsequent tests, otherwise it is ignored. If a human supervisor is available, he is asked to label the sample so it can be added to the memory. The average identification performance obtained so far on box-like objects of the same color is of 83%.

As an example, Fig. 2 shows object identity plotted as a function of SOS and AOS activations, and of estimated height. While for some objects these values are very informative and nearly enough to recognition, it is apparent that other objects cannot be resolved from each other only using one or two visual features. If we should decide to include grasping in our experiments, object weight, estimated upon lifting thanks to haptic perception of the grasped object, can also be used to distinguish and recognize objects.

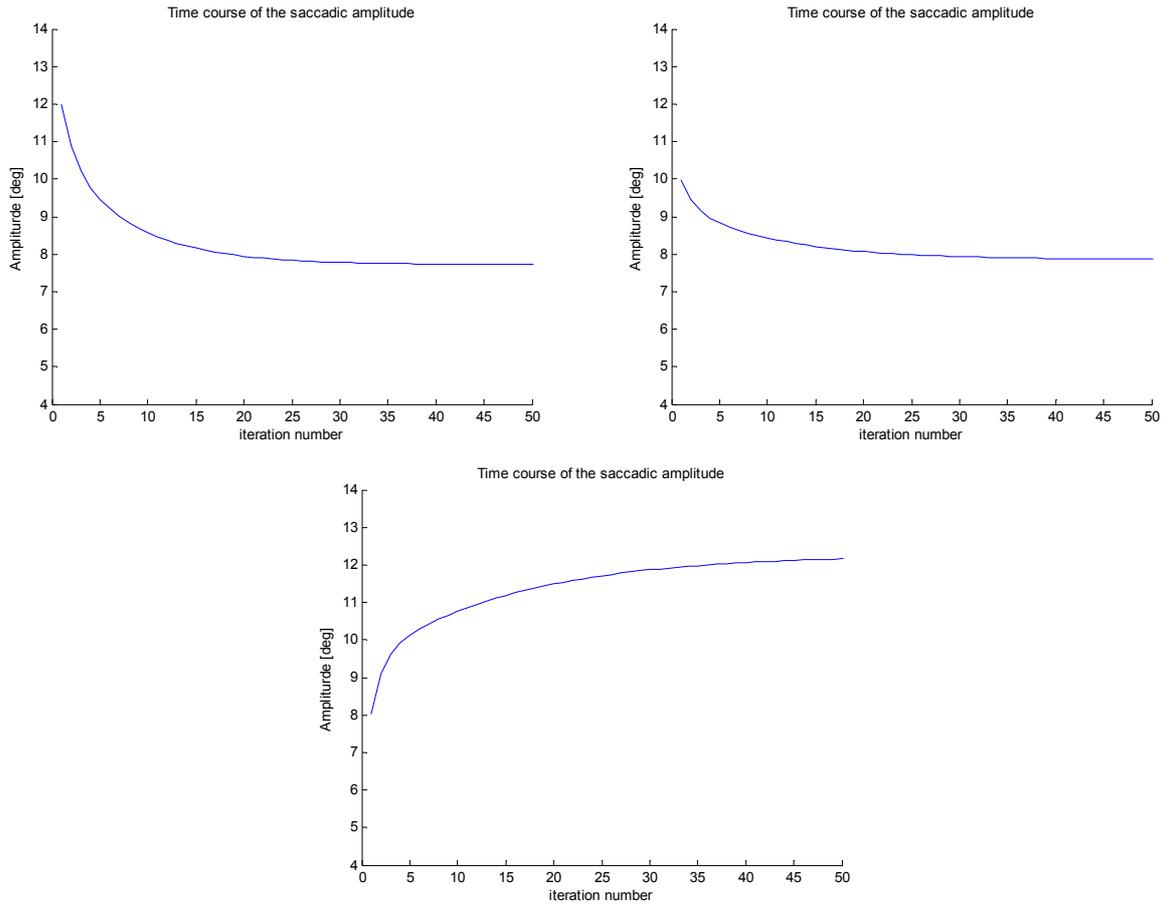
#### 4. Reproduction of psychophysical findings – saccadic adaptation

Regarding the psychophysical effects the model and its robotic implementation are expected to be able to reproduce, we are beginning to check the model behavior in the case of the deceptive visual feedback, such as in typical experiments of saccadic adaptation. This is done by eliciting a saccade (based on vergence/version eye movement control) toward a given visual target, and providing a fictitious error on the reached final position. For the computational model, this is achieved by adding an offset to the output. On the robot, the same effect will be obtained moving the visual target as for human subjects. For this purpose, we implemented an experimental setup similar to that used in the experimental protocols of human saccadic adaptation experiments.

Analysis of how (as in the saccadic adaptation protocol) such artificial displacement of the target affects the artificial agent oculomotor and arm motor abilities can serve as a validation of the underlying model, and may help in advancing hypotheses on saccadic adaptation mechanisms in humans and monkeys. We were able to verify that our model actually exhibits saccadic adaptation, altering its ability to perform correct saccades according to the deceptive feedback. We are performing a number of different experiments and the analysis of error distributions around the target point and of error vectors is providing interesting information that we are studying together with the partner WWU. Some examples of the obtained results, depicted in Fig. 3 and 4, are described below.

Once the network described in Section 3 has been trained, the process of saccadic adaptation can start. The learning strategy which has been used for training the net during the altered perception stage is the delta rule. In this phase, the system is ideally fixating a starting point at  $(x_F, y_F)$  and a target  $(x_0, y_0)$  that triggers an input  $(C_{x_0}, C_{y_0}, D_0)$  in the cyclopean, disparity space is shown. This input produces an output  $(V_{x_0}, V_{y_0}, V_{s_0})$  composed of the version and vergence eye motor components necessary for gazing at the target. The vergence and version angles which are estimated by the net will be incorrect, because in the meantime, the target has been displaced to the point  $(x_1, y_1)$ , correspondent to a different visual input and a different transformation  $(V_{x_1}, V_{y_1}, V_{s_1})$ . The network then changes the weights based on the error movement it committed:  $(V_{x_1}-V_{x_0}, V_{y_1}-V_{y_0}, V_{s_1}-V_{s_0})$ . As suggested by human studies (Wallman & Fuchs, 1998), a saccade to the new position is not required, and the actual correct position is estimated from retinal information by the network itself. After each deceptive feedback movement, a successful gaze movement back to the fixation point is simulated. No error is introduced in this case. The saccadic movement is repeated 100 times and each iteration allows to reduce the saccadic error due to target displacement, thus simulating saccadic adaptation.

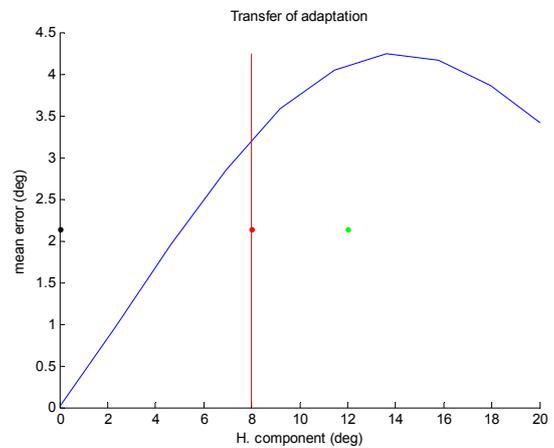
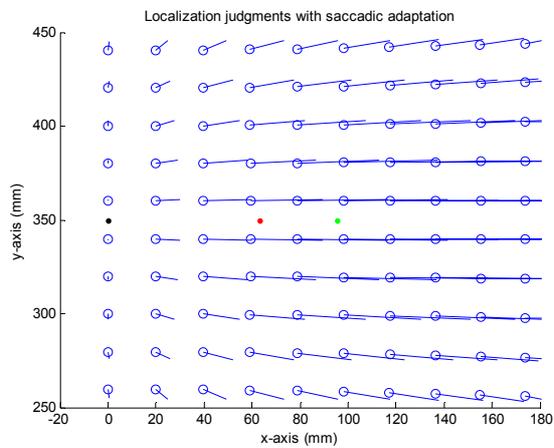
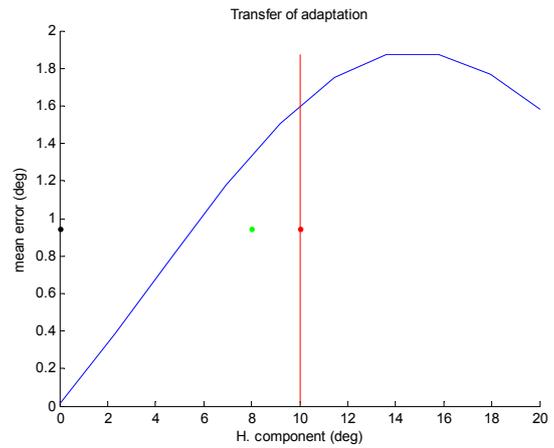
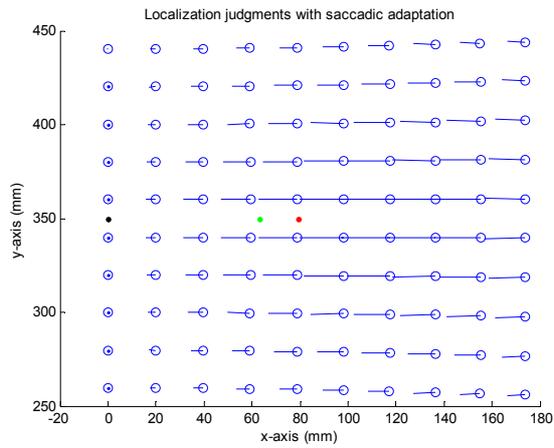
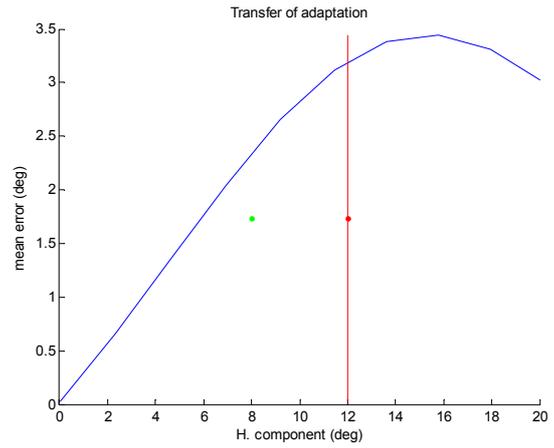
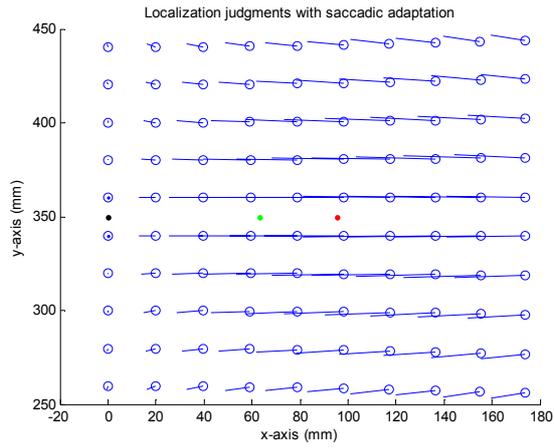
In the three examples we show, there is a horizontal displacement of the visual target which deceive the visual feedback after movement execution as described above. In the first two cases, the target is placed at  $(x, y) = (12, 0)$  and  $(10, 0)$  in degrees, respectively, and then displaced inward at  $(8, 0)$ , to check the effect of different displacement amplitudes. In the third case the displacement is in the opposite, outward direction, being the target at  $(8, 0)$  and the displaced feedback at  $(12, 0)$ . Fig. 3 plots the learning curves in which the system adapts to the deceptive feedback and learns the new gaze movement to the displaced target in a few dozen of iterations.



**Figure 3:** Learning curves of the saccadic adaptation experiments, in which a new gaze movement is learned.

After 100 iterations, when the system has fully adapted to the deceptive feedback, the response of the network with the new weights is compared to that with initial weights. This comparison takes place for different locations in space. We show the comparison done on a grid of points on the space  $(x, y)$  on the left side of Fig. 4, for the same experimental conditions of Fig. 3. It can be observed how the altered response transfers to different values of  $x$ , inward and outward respectively, and even on the  $y$  coordinate, especially for the larger displacements of the first and third experimental conditions. These results resemble those observed in human saccadic adaptation studies, such as Collins *et al.* 2007 and Schnier *et al.* (submitted).

A clearer assessment of how saccadic adaptation transfers to values of horizontal displacement different from what the system was trained to is depicted in the right side of Fig. 4, where the error is plotted as a function of the  $x$  component, cumulated across the values of  $y$  depicted on the left side graphs. As revealed by the experiments described in Collins *et al.* 2007, there is peak in the saccadic transfer error external to the target for the inward adaptation scheme. A similar peak is seen for outward adaptation in the model simulations. However, human outward adaptation appears to be less peaked and more uniform than inward adaptation (Schnier *et al.*, submitted), suggesting that inward adaptation and outward adaptation use partly different learning mechanisms.



**Figure 4:** LEFT: Displacement errors after saccadic adaptation in the (x, y) space. Black dot: fixation point; red dot: target; green dot: displaced feedback. RIGHT: Cumulated adaptation transfer on the horizontal direction; red dot: target; green dot: displaced feedback.

## **5. Stream integration – sensorimotor memory of nearby objects**

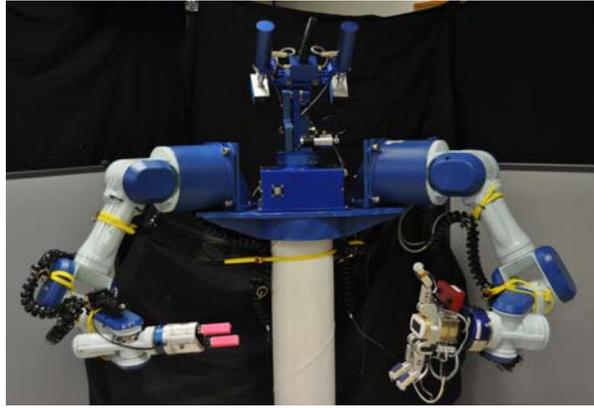
At this stage, we need to introduce recognition and memory in the sensorimotor interaction schema. This is done implementing an actual integration with object recognition modules. Such integration is being pursued through a sequence of increasingly demanding experimental conditions. At the current stage of development, integration is being performed in the experimental environment resembling psychophysical experiments performed in WP5, in which stimuli are constituted by colored lights on a screen. Object recognition is thus very easy at this point, but the fundamental principles for contextually encoding object identity and sensorimotor parameters required to gaze/reach at it are already included in the schema. Later on, acting on more complex and real-life environments, more complex object recognition methods will be employed, such as the ventral stream modeling described in Section 4 and the attention based identification system developed by the Chemnitz partner, with lower-level biological inspiration.

The behavioral schema in which object identity is associated to a particular sensorimotor configuration starts when the system is already able to accurately gaze, and possibly reach, toward a visual target. In previous experiments, this was done with the presence of only one stimulus in the visual field of the robot. Now, more targets can be visualized contextually. In this associative learning schema the robot is gazing at a given fixation point while two or more stimuli appear in its field of view. The robot is thus required to use its visuomotor skills, embedded in the neural networks that transform between visual and motor parameters, to estimate the eye motor movements required to fixate on each visible object, without actually executing the saccadic movement on the targets. The sums of the movement vectors with the actual gazing direction constitute an instance of the absolute positions of the visible objects. Next, the fixation point is changed, the robot gazes at the new one and estimates the new movement vectors required to fixate on the visual targets, creating new instances of the targets absolute position. The process is then repeated and at each step a slightly different absolute position is computed for each visual target. Ideally, all computed positions are the same, but due to unavoidable distortions and imprecisions we expect a range of variability, and the average of each estimated location is stored as the memory of the visual target position, for all targets. After the learning process, the system can be required to fixate and/or reach a target given its identity, using the visuomotor associations it has learned during the previous step. Ideally, this can be done even on objects placed out of the field of view, if their location has been previously observed.

## **6. Hardware setup and setting of robotic experiments.**

One of the main goals of our research is to provide the robot with advanced skills in its interaction with the environment, namely in the purposeful exploration of the peripersonal space and the contextual coding and control of eye and arm movements. The implementation on an actual, physical sensorimotor setup is a potential source of additional insights for the computational model, hardly achievable with simulated data.

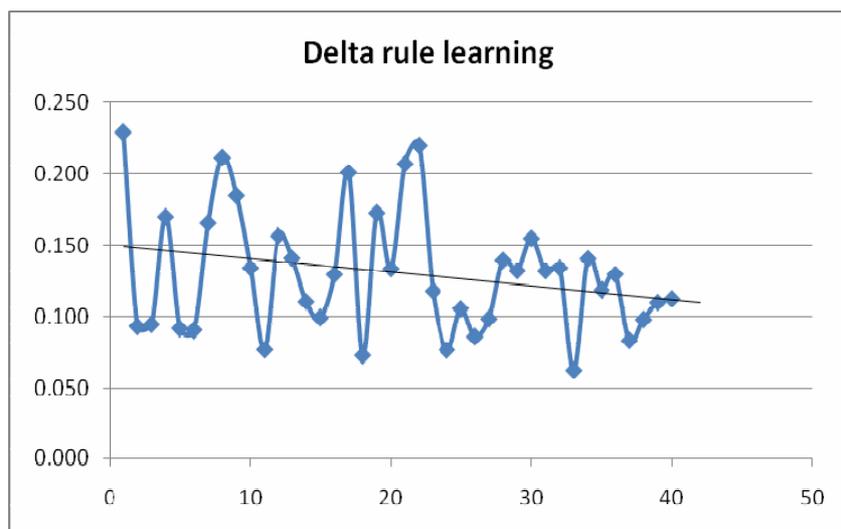
Our robotic setup consists of a torso body which mounts a pan-tilt-vergence stereo head, and two multi-joint arms, endowed with a parallel-jaw gripper and a three finger Barrett Hand respectively (see Fig. 5). Visual, visuomotor and arm motor libraries have been developed and adapted to this setup. Among other utilities, we developed a library, which can be accessed by software or as a user interface, for the coordinated control of head movements according to a vergence/version control, as in Eyeshots objectives, and a second one with several visual processing steps to assess the suitability of various algorithms and parameters in different conditions and visual setups.



**Figure 5:** Humanoid robot with pan/tilt/vergence head and arm with hand.

Before applying the model to the humanoid robot we made an effort to include already a full 3D functionality (the model was originally in 2D, allowing only for movements in the  $xz$  space). We thus introduced tilt movements of the head and one more joint for the arm. For simplicity, and waiting for the oncoming integration with specialized modules from the partners, at the moment visual stimuli appearing on a computer screen as in WP5 experiments are being used. As a first step, aimed at configuring the visual/oculomotor network on the robot, a stimulus appears on the screen, the robot estimates the oculomotor movement required to gaze at it and execute it. The residual distance of the stimulus from the center of the image is used to estimate the movement error, which is employed to adjust the network weights through incremental learning by delta rule. In this way, the exploration behavior is aimed at learning an implicit sensorimotor map of the environment. This is not done from scratch, as learning is bootstrapped with the weights learned during the training of the modeled network. The robot learns to adjust the weight to its actual parameters and to any distortions of image and kinematics movements. A preliminary result is the learning curve depicted in Fig. 6, obtained in uncontrolled experimental conditions with an object placed on a table. Full experimentation of the learning framework and on the computer monitor setup is in progress.

In the second learning stage, in which arm movements are introduced, possible misalignments are made of two different error components, one due to the visual-oculomotor transformation and the other to the arm-oculomotor. The two error components can be estimated measuring the visual distance between the effector and the final gazing point. A marker is placed on the forefinger to discriminate between oculomotor and arm-motor errors. The use of tactile feedback is planned but not applied yet for safety reasons. Again, extensive experimentation of this learning framework is under development.



**Figure 6:** Error (mm) of robot gazing movements toward visual stimuli while the delta rule is applied to adjust the weights of the visual/oculomotor network.

## References

Chinellato, E., Antonelli, M., Grzyb, B. & del Pobil A. P. (2010) Implicit sensorimotor mapping of the peripersonal space by gazing and reaching. *IEEE Transactions on Autonomous Mental Development*, In Press.

Chinellato, E. & del Pobil, A.P. (2008) Neural Coding in the Dorsal Visual Stream. *From Animals to Animats*, LNAI 5040, pp. 230-239.

Collins, T., Dore-Mazars, K., and Lappe, M. (2007) Motor space structures perceptual space: Evidence from human saccadic adaptation. *Brain Res.*, 1172:32-39.

Schnier, F., Zimmermann, E. and Lappe, M. (submitted) Adaptation and mislocalization fields for saccadic outward adaptation in humans.

Wallman, J. & Fuchs, A. (1998). Saccadic gain modification: Visual error drives motor adaptation. *J. Neurophysiol.*, 80: 2405-2416